

10. De kwaliteit van SERVQUAL

een toepassing bij zeven dienstverlenende organisaties

A. SMIDTS

SAMENVATTING

Het SERVQUAL-model met bijbehorende standaardvragenlijst is een hulpmiddel bij de diagnose van kwaliteitsproblemen in dienstverlening. De methode mag zich in een groeiende belangstelling verheugen bij dienstverleners in de praktijk. In dit artikel worden de resultaten vergeleken van een toepassing van de methode bij een zevental dienstverlenende organisaties. Hierdoor wordt inzicht verkregen in de betrouwbaarheid en validiteit van de methode en in de praktische bruikbaarheid. Daarnaast wordt met deze vergelijking een referentiekader geboden voor de grootte van de verschillende gaps in het model. Op deze wijze kan een organisatie haar eigen resultaten met de methode afzetten tegen onze resultaten bij de zeven organisaties. Uit de bevindingen blijkt dat het model op veel punten een betrouwbaar en valide beeld geeft van de externe kwaliteitstekorten en de interne oorzaken daarvoor. Verbeteringen in de methode kunnen met name aangebracht worden bij: de meting van de verwachtingen (biedt te weinig extra informatie), de meting van de interne organisatiegaps (weinig bruikbaar want erg globaal) en de meting van de antecedente factoren voor de interne gaps (te onbetrouwbaar). Concrete aanbevelingen worden gedaan om de waarde van de methode te verhogen.

1. INLEIDING¹

De kwaliteit van dienstverlening staat al weer enige jaren in grote belangstelling van zowel de wetenschap als de praktijk. In de marketing heeft het SERVQUAL-model van Parasuraman, Zeithaml en Berry (1985, 1988) verreweg de meeste aandacht gekregen. Eén verklaring voor de aandacht kan zijn dat het model relatief eenvoudig en intuïtief aansprekend is. Het model, waaraan sinds 1983 gewerkt is, bestaat uit enkele diagrammen waarin de meest belangrijke variabelen die kwaliteitsoordelen van klanten bepalen geïnventariseerd zijn. Een tweede aantrekkelijk aspect van het model betreft de verbinding die gelegd wordt tussen de externe beoordeling van kwaliteit door klanten en een aantal interne factoren in de organisatie. Een derde succesfactor is waarschijnlijk de standaardvragenlijst die de auteurs bij het model geleverd hebben. Dit maakt SERVQUAL tot een gemakkelijk toe te passen hulpmiddel bij de diagnose van kwaliteitsproblemen.

Inmiddels heeft niet alleen de marketingwetenschap, maar ook de praktijk het model en de vragenlijst ontdekt. Het is in korte tijd een populaire methode geworden om de kwaliteit van dienstverlening in kaart te brengen. Ook in Nederland is volgens Lemmink (1992) al een groot aantal empirische onderzoeken uitgevoerd met het model en uit eigen observatie blijkt dat in menige bedrijfskundige afstudeerscriptie die betrekking

heeft op dienstverlenende organisaties, de methode wordt toegepast. Hiermee lijkt het model en het meetinstrument in de groeifase van de levenscyclus te verkeren. De keerzijde van het succes is echter dat er enige sleet in de methode kan komen. De indruk bestaat dat de toepassingen soms iets te onnadenkend uitgevoerd worden en dat SERVQUAL dreigt te reduceren tot een standaardtruc. Daarnaast heeft het model enigszins te kampen met te hoog gespannen verwachtingen, alsof het model het medicijn is voor alle kwaliteitskwalen.

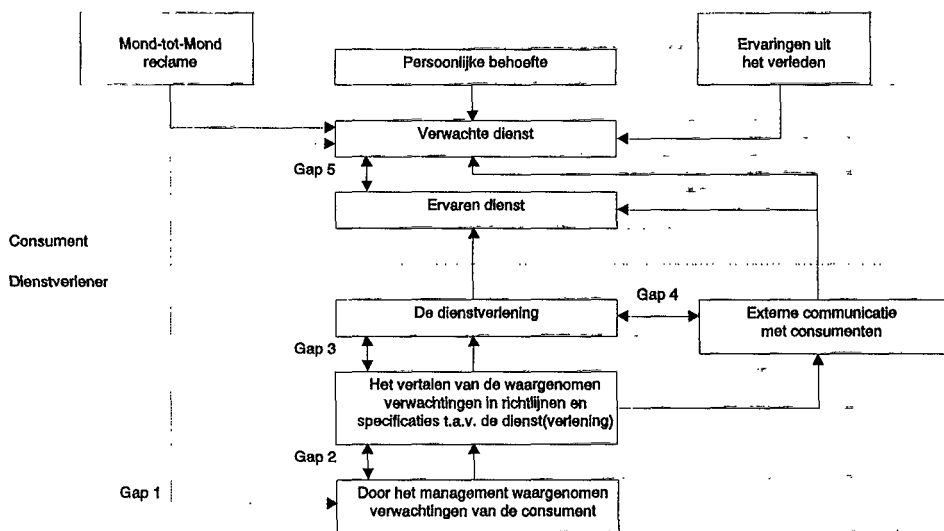
De doelstelling van dit artikel is het bespreken van de resultaten van en de ervaringen met een aantal toepassingen van SERVQUAL in de praktijk. Niet alleen willen we op deze wijze meer inzicht krijgen in de validiteit en betrouwbaarheid van het model en de methode, maar ook komen met aanbevelingen over hoe de methode het best toegepast kan worden en welke risico's het gebruik van de methode heeft. Daarnaast kan dit overzicht helpen bij het scheppen van een referentiekader voor organisaties die de methode gebruiken. Bij iedere toepassing komt namelijk vroeg of laat de vraag aan de orde of een bepaalde meting in de eigen organisatie 'groot' of 'klein' is in vergelijking met die bij andere organisaties.

Opgemerkt moet worden dat het in dit artikel uitdrukkelijk niet de bedoeling is uitgebreide, fundamentele kritiek op het model te geven. Hiermee wil niet gezegd zijn dat deze kritiek niet mogelijk is of niet relevant is. Vele anderen hebben zich reeds in kritische zin over het model uitgelaten en ook de ontwerpers van het model blijven werken aan verbeteringen. Zo is onlangs een belangrijke theoretische verdieping gegeven aan de verwachtingencomponent van het model (Zeithaml et al. 1993) en is het statische model omgezet in een dynamisch model (Boulding et al. 1993). Wij beperken ons hier echter tot een kritische blik op een standaardtoepassing van het oorspronkelijke SERVQUAL-model en stellen ons de vraag wat van die toepassingen geleerd kan worden in afwachting van fundamentele verbeteringen. Hiertoe worden de resultaten bij zeven dienstverlenende organisaties met elkaar vergeleken en wordt het model op metaniveau beoordeeld.

2. HET SERVQUAL-MODEL

Voor een beschrijving van het model en de methode wordt verwezen naar de oorspronkelijke auteurs. Met name het boek van Zeithaml et al. (1990) geeft een uitstekende inleiding. Een uitgebreide Nederlandstalige beschrijving is recent verschenen in dit Jaarboek (Lemmink 1992). Op deze plaats wordt de bespreking van het model zeer beknopt gehouden.

De essentie van het model is dat kwaliteit gedefinieerd wordt als het verschil tussen de dienstverlening die de klant *verwacht* en de dienstverlening die de klant *ervaart* (percipeert) (zie figuur 1). Wanneer de dienstverlening niet tegemoet komt aan de verwachting is er sprake van een kwaliteitstekort (een gap). De kwaliteit van dienstverlening wordt volgens het model gebaseerd op vijf kwaliteitsdimensies: (1) de tastbare aspecten van de dienst en dienstverlener ('tangibles'); (2) de betrouwbaarheid: 'nauwgezet doen wat je belooft'; (3) de responsiviteit: 'echte bereidheid om de klant te helpen en deze prompt en alert van dienst te zijn'; (4) de deskundigheid en beleefdheid ('assurance'): 'de geloofwaardigheid, deskundigheid en hoffelijkheid van het personeel; de klant het vertrouwen geven dat hij/zij in goede handen is'; en (5) de empathie: 'begrip en aandacht voor de klant als een individu en zijn/haar specifieke wensen'.²



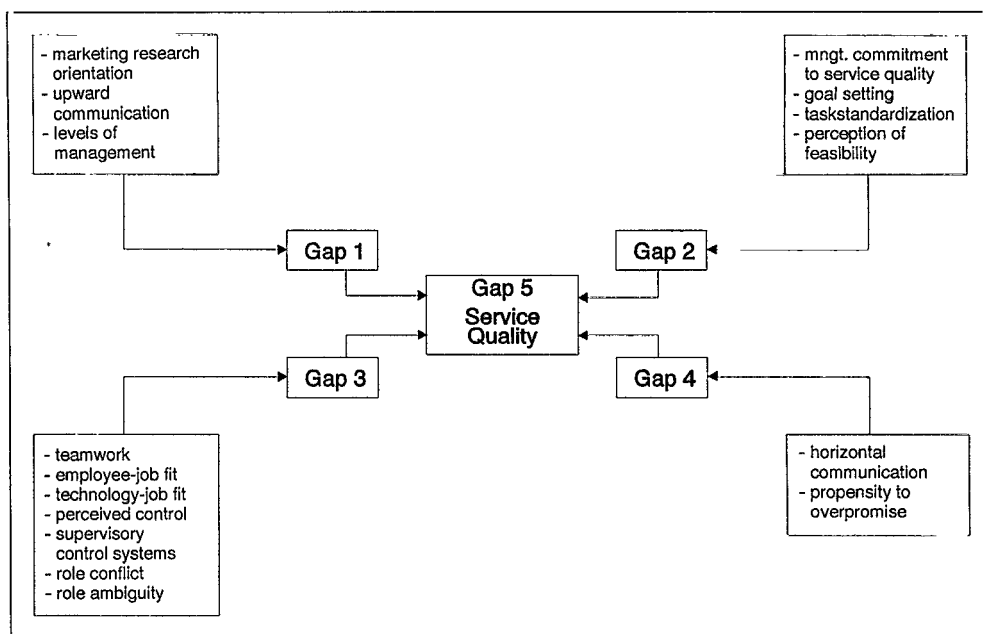
Figuur 1. Het SERVQUAL-model.

Bron: Zeithaml et al. (1990).

Ieder van de vijf dimensies wordt gemeten met behulp van een aantal attributen (items). In totaal zijn er in de standaardvragenlijst 22 items geformuleerd om de kwaliteit in kaart te brengen. Naast de perceptie en de verwachtingen met betrekking tot de 22 items, wordt klanten ook gevraagd aan te geven wat het relatieve belang is van de vijf kwaliteitsdimensies. Hiertoe dient de respondent 100 punten te verdelen over de vijf dimensies naar belangrijkheid.

De grootte van de kwaliteitsgap (gap 5) wordt in het model afhankelijk gesteld van de grootte van een viertal interne organisatiegaps. Als klanten een tekort aan kwaliteit ervaren dan kan dat komen doordat: (a) het (top)management slecht op de hoogte is van wat klanten eigenlijk verwachten van de dienstverlening (gap 1); (b) het (top)management er onvoldoende in slaagt de verwachtingen van klanten vast te leggen in expliciete richtlijnen en specificaties voor de dienstverlening (gap 2); (c) de medewerkers er onvoldoende, of niet consistent, in slagen de dienstverlening op het gewenste niveau te leveren (gap 3), en (d) de externe communicatie niet aansluit bij de in werkelijkheid geleverde dienstverlening (het niet volledig kunnen voldoen aan de beloften). De grootte van gaps 1 en 2 wordt gemeten bij het (top)management, die van gaps 3 en 4 bij de medewerkers die in direct contact staan met de klant.

In het zogenaamde Extended-SERVQUAL model is de grootte van ieder van de vier gaps afhankelijk gesteld van een aantal antecedente factoren. Deze antecedenten zijn in figuur 2 opgesomd. Hierbij is voor de duidelijkheid de Engelstalige benaming gehandhaafd. In een enkel voorbeeld worden deze antecedente factoren hier toegelicht. De grootte van gap 1 (het management is onvoldoende bekend met de verwachtingen van



Figuur 2. De antecedente factoren voor de vier interne kwaliteitsgaps.

Bron: Zeithaml et al. (1990).

klanten) wordt volgens het model vooral bepaald door: (1) te weinig marktonderzoek onder klanten, (2) een slechte 'upward communication' zodat kennis van klanten die bij medewerkers op de werkvloer aanwezig is, niet aankomt op (top)management niveau, en (3) teveel lagen in de organisatie die de informatiestroom van beneden naar boven belemmeren en de afstand van management tot medewerker en klant erg groot maakt. Voor de overige gaps gelden vergelijkbare factoren.

3. DATAVERZAMELING

In dit artikel worden de resultaten besproken van de toepassing van het SERVQUAL-model bij zeven dienstverlenende organisaties. Deze organisaties zijn: twee concurrerende *vakantiebungalowparken*, een *landelijke vereniging* (professioneel georganiseerd), een *regionale luchthaven*, een *touroperator*, een *restaurantketen* en een *ziekenhuis*. De meeste, maar niet alle organisaties, kunnen beschouwd worden als 'A-merken' in hun branche. De organisaties zijn divers qua dienstverlening. Tegenover de bungalowparken waar de fysieke elementen van de dienstverlening zoals de bungalow en het zwembad heel belangrijk zijn, staan bijvoorbeeld de touroperator en de vereniging waar de tastbare aspecten van de dienstverlening relatief minder relevant zijn. Naast diensten die de consument goed kan beoordelen op kwaliteit (bijvoorbeeld het restaurant) staat de dienstverlening in het ziekenhuis waar vertrouwen in de kwaliteit voorop staat. Tevens zijn twee niet-op-winstgerichte organisaties (vereniging en ziekenhuis) in het onderzoek opgenomen. Het samenstel van organisaties maakt het moge-

lijk tot op zekere hoogte de validiteit van SERVQUAL te bepalen. Zo zou het bijvoorbeeld verwondering wekken als de 'tangibles' dimensie bij bungalowparken door klanten niet belangrijker wordt gevonden dan dezelfde dimensie bij de touroperator of de vereniging.

Vanwege de gewenste geheimhouding van de individuele organisaties zullen de resultaten in dit artikel op totaalniveau gerapporteerd worden. Meestal zal naast minimum en maximum waarden van een variabele, het gemiddelde of de mediaan van de zeven organisaties gegeven worden. Bij de interpretatie en toelichting zal gewezen worden op specifieke resultaten voor bepaalde organisaties, meestal zonder de precieze scores te vermelden.

Voor alle organisaties is zowel het interne deel als het klantendeel van SERVQUAL uitgevoerd. Er is een Nederlandse vertaling gemaakt van de vragenlijst die is gepubliceerd in Zeithaml et al. (1990). Deze vertaling was overigens niet altijd eenvoudig en is van te voren uitvoerig getest bij verschillende proefpersonen en vervolgens aangepast. Om een optimale vergelijking tussen de organisaties te kunnen maken, zijn zo weinig mogelijk organisatiespecifieke veranderingen in de vragenlijst aangebracht. Sporadisch zijn vragen aan de lijst toegevoegd. Dit is met name het geval geweest bij de bungalowparken waar de 'tangibles' dimensie is uitgebreid met extra vragen over de kenmerkende fysieke elementen van het park. Een enkele keer zijn enige items verwijderd of veranderd omdat het aspect bij de specifieke organisatie niet van belang bleek of een andere inhoud had.

Alle vragenlijsten zijn schriftelijk afgenomen. Bij de touroperator en het ziekenhuis is een postale enquête gehouden, bij de overige organisaties werden de vragenlijsten na afloop van de dienstverlening ter plekke door klanten ingevuld. De dataverzameling werd uitgevoerd door derde- en vierdejaars studenten bedrijfskunde. Het veldwerk vond plaats in 1991.

In tabel 1 worden de aantallen respondenten en responspercentages vermeld. Het minimum aantal klanten dat is ondervraagd bedraagt 111 en het maximum aantal 1789. Dit laatste aantal is uitzonderlijk (het op één na hoogste aantal respondenten bedraagt 241) en heeft betrekking op een uitgebreid onderzoek bij één van de bungalowparken. In twee verschillende vestigingen werden zowel midweekgasten als weekendgasten ondervraagd. Naast Nederlandse gasten werd ook de mening gevraagd van Belgische en Duitse gasten. Hierdoor was het mogelijk zeer specifieke verschillen tussen marktsegmenten en vestigingen op te sporen met betrekking kwaliteitsoordelen en verwachtingen. De responspercentages lopen uiteen van 37% (bij het ziekenhuis) tot 87% (op de luchthaven). Op de luchthaven zijn overigens zakenreizigers geïnterviewd (business class reizigers, waarvan de helft Nederlanders) die zeer bereid bleken de vragenlijst in te vullen.

De vragenlijsten voor de medewerkers en het (top)management zijn ook schriftelijk afgenomen. Bij het management waren de interviewers soms persoonlijk aanwezig tijdens het invullen. Hierdoor ontstond een gecombineerd gesloten en open interview. Bij diverse vragen kwam het management met extra opmerkingen naar aanleiding van een vraag. Dit gaf vaak waardevolle informatie die bij de interpretatie van de resultaten is gebruikt. Ook bij de medewerkers zijn interviewers geregeld aanwezig geweest om opmerkingen en anekdotes te noteren. De indruk bestaat dat deze combinatie van een gestructureerde vragenlijst met de mogelijkheid tot doorvragen, zeer aanbevelingswaardig is.

Tabel 1. Responsaantal (n) en responspercentage (%) bij zeven organisaties.

		Mediaan	Minimum	Maximum
Klanten	n	149	111	1789
	%	53%	37%	87%
Medewerkers	n	47	15	156
	%	46%	25%	89%
(Top)management	n	14	7	24
	%	88%	28%	100%

Het responspercentage bij de medewerkers lag bij vijf van de zeven organisaties iets onder de 50% (zie tabel 1). De laagste respons was bij de vereniging (25%). Een gunstige uitzondering was de touroperator met een responspercentage van 89%. Deze organisatie heeft zich veel moeite getroost om voor medewerkers het belang van het onderzoek aan te geven. Voor een hoge respons is het vooral zaak absolute anonimiteit te garanderen en de medewerkers daarvan ook te overtuigen. Met name het deel van de vragenlijst dat betrekking heeft op de antecedente factoren voor de interne gaps, bevat vragen die te maken maken hebben met het oordeel van de medewerker over het management. Het bieden van zekerheid aan respondenten dat men vrijuit kunnen praten zonder consequenties, is dan uiteraard zeer wezenlijk. Bij de interpretatie van de gegevens moet rekening gehouden worden met de mogelijkheid dat een deel van de non-respons bestaat uit zelfselectie. Mogelijk hebben vooral tevreden medewerkers de vragenlijst ingevuld. Hierdoor kunnen sommige resultaten te positief gekleurd zijn.

De betrokkenheid van het management bij het onderzoek lijkt in het algemeen hoger dan van de medewerkers. De mediaan van de respons ligt bij 88% en bij een tweetal organisaties was de respons zelfs 100%. In het ziekenhuis bleek de respons vrij laag (27%). Een mogelijke verklaring voor deze teleurstellende respons kan zijn dat een aantal hoofden van afdelingen kwaliteit vooral associeert met medisch-technische kwaliteit en de klantgerichte definitie van kwaliteit, zoals die in SERVQUAL naar voren komt, minder relevant vindt.

Onze ervaringen bij de zeven organisaties laten zien dat bij een SERVQUAL toepassing rekening gehouden kan worden met een respons van ongeveer 50% bij klanten en medewerkers en een vrij hoge respons bij (top)management. Bij voorkeur dienen de schriftelijke vragenlijsten bij ieder van de drie groepen respondenten aangevuld te worden met een aantal open interviews naar aanleiding van de vragenlijst. Op deze wijze kan extra informatie verzameld worden die van groot belang blijkt bij de interpretatie en rapportage van de 'kale' cijfers.

4. HET OORDEEL VAN KLANTEN OVER DE KWALITEIT

4.1 Het relatieve belang van de kwaliteitsdimensies

In tabel 2 zijn, zowel voor klanten als voor het management, de gemiddelde gewichten weergegeven die aan de kwaliteitsdimensies worden gehecht. De gewichten sommeren tot vijf hetgeen wil zeggen dat als alle dimensies even belangrijk zijn, elke dimensie het

gewicht van 1 heeft. Per dimensie zijn tevens minimum en maximum waarden vermeld.

Tabel 2. Belangrijkheid van de vijf dimensies volgens klanten en volgens (top)management (tussen haakjes de rangorde met 1 = het belangrijkste).

<i>Dimensies</i>	<i>Gemiddelde</i>		<i>Minimum</i>		<i>Maximum</i>	
	Klanten	Mngt.	Klant	Mngt.	Klant	Mngt.
Tastbare zaken	1.00 (2)	0.99 (2)	.65	.43	1.55	1.95
Betrouwbaarheid	1.34 (1)	1.32 (1)	.90	.65	1.80	2.08
Responsiviteit	0.97 (3)	0.98 (3)	.86	.88	1.00	1.21
Deskundigheid	0.91 (4)	0.87 (4)	.75	.65	1.10	1.05
Empathie	0.78 (5)	0.84 (5)	.65	.70	.95	.96

Als voor klanten de gemiddelde gewichten vergeleken worden dan blijkt dat betrouwbaarheid de belangrijkste dimensie wordt gevonden (1.34). Deze bevinding blijkt vrij universeel te zijn (Zeithaml et al. 1990). De tastbaarheden dimensie komt op de tweede plaats. Empathie wordt gemiddeld bij deze zeven organisaties het minst belangrijk gevonden als kwaliteitsdimensie (0.78); deze dimensie komt telkens op de vierde of de vijfde plaats qua belangrijkheid.

Het algemene patroon dient echter sterk genuanceerd te worden. Het belang van de ‘tangibles’ varieert maximaal tussen de organisaties. Bij drie organisaties is het de belangrijkste dimensie: bij de twee bungalowparken (gemiddeld 1.5) en bij de restaurantketen (1.2), bij de overige organisaties is het de minst belangrijke dimensie (gemiddeld 0.7). Betrouwbaarheid scoort consistent hoog (als eerste of tweede). Responsiviteit is vaak de tweede of derde dimensie. Opvallend is dat empathie, de individuele en zorgzame aandacht voor de klant, bij de klanten op de vierde of vijfde plaats komt behalve bij de patiënten in het ziekenhuis. Patiënten vinden empathie van groot belang bij de beoordeling van kwaliteit (een gedeelte tweede plaats met responsiviteit, 0.95). Interessant is dan om na te gaan in hoeverre het ziekenhuis er ook in slaagt deze empathie te bieden aan de patiënten. Hier komen we later op terug.

Ook aan het management is gevraagd aan te geven hoe belangrijk de klanten de dimensies vinden. Het idee hierbij is dat als het management denkt dat klanten andere zaken belangrijk vinden dan wat klanten werkelijk vinden, zij zich bij het marketingbeleid op de verkeerde dimensies concentreren. Gemiddeld blijken management en klanten tot dezelfde rangorde van de dimensies te komen. Ook per organisatie komt de volgorde redelijk overeen. Er zijn echter een aantal opmerkelijke verschillen te constateren. Bij één van de bungalowparken blijkt dat het management denkt dat responsiviteit aanzienlijk belangrijker wordt gevonden dan de klanten zelf aangeven (1.2 tegenover 1.0), terwijl het belang van de ‘tangibles’ door het management wordt onderschat (1.2 tegenover 1.4). Het management van het andere bungalowpark overschat het belang van de ‘tangibles’ echter sterk (2.0 tegenover 1.6). Hoewel bij beide parken de klanten weinig verschillen in het belang dat zij aangeven, verschilt het management mogelijk wel in de nadruk die zij leggen op de fysieke omgeving als bepalend voor het kwaliteitsoordeel.

Dit verschil zou tot uiting kunnen komen in de gekozen marketingstrategie. Ook bij het ziekenhuis is de rangorde van belangrijkheid gelijk tussen management en patiënten. Opvallend is echter de grote nadruk die het management legt op betrouwbaarheid. Met 2.1 is dit de hoogste score van de zeven organisaties, terwijl patiënten het belang van betrouwbaarheid aangeven met 1.4. Daartegenover onderschat het management het belang dat patiënten in een ziekenhuis hechten aan de 'tangibles' (0.4 tegenover 0.8). Een mogelijke verklaring voor deze verschillen kan zijn dat het management de hotelfunctie van een ziekenhuis voor een klant onderschat. Terwijl het management bij 'tangibles' vooral gefocussed zal zijn op moderniteit van de medisch-technische apparatuur, gaan patiënten er wellicht vanuit dat die apparatuur per definitie perfect is en houden zij bij de beoordeling van de kwaliteit relatief meer rekening met de 'tangibles' die te maken hebben met de hotelfunctie van een ziekenhuis.

4.2 De verwachtingen en percepties van klanten

De resultaten in deze paragraaf zullen vermeld worden op dimensieniveau. Uiteraard is het zo dat voor de organisaties de rapportage niet beperkt is gebleven tot deze samengevatte informatie. Juist de resultaten op itemniveau bieden, naast de gemiddelden op de dimensies, zeer veel detailinformatie over waar de kwaliteitsproblemen zitten. Zo is het percentage ontevreden per aspect vaak zeer instructief.

In tabel 3 zijn de gemiddelde verwachtingen en percepties weergegeven, naast per dimensie het minimum en het maximum. De verschillen tussen verwachtingen en percepties is de meting van de kwaliteit (gap 5 in het model). Het blijkt dat de verwachtingen van klanten hooggespannen zijn: gemiddeld 6.27 op een schaal van 1 t/m 7. De hoogste verwachtingen hebben de klanten van het ziekenhuis (gemiddeld 6.6), de laagste verwachtingen komen voor bij de restaurantketen (6.0). De gemiddelde perceptie van de kwaliteit is echter ook hoog (5.81), met als resultaat dat het kwaliteitstekort beperkt is (gap 5 bedraagt 0.46).

Vergeleken met de andere organisaties voldoet de luchthaven het minst aan de verwachtingen. De kwaliteitgap is hier het grootst (1.0). Dit komt vooral doordat de beoordeling vrij laag is (5.1), terwijl de verwachtingen realistisch zijn (6.1). Mogelijk wordt de lage beoordeling veroorzaakt door het feit dat zakenreizigers zeer kritisch staan tegenover dienstverlening en over veel vergelijkende ervaring met andere luchthavens beschikken. Het zou in dit verband interessant geweest zijn naast zakenreizigers ook tourist-class reizigers te interviewen om deze hypothese te onderzoeken.

De op één na grootste kwaliteitgap blijkt te bestaan bij het ziekenhuis. Dit komt echter niet omdat patiënten de kwaliteit als laag percipiëren. Integendeel, het ziekenhuis scoort zeer hoog. De reden is echter dat de verwachtingen zeer hooggespannen zijn met als gevolg dat de gap relatief groot is (0.54). Overigens is de perceptie van de bij het ziekenhuis belangrijke dimensie empathie het hoogst van alle organisaties (6.3), maar valt het niet mee te voldoen aan de hoge verwachtingen (6.7).

Als we de kwaliteitgap vergelijken tussen de dimensies dan blijkt de grootste gap te bestaan voor de belangrijkste dimensie: betrouwbaarheid (0.68). In de literatuur is er reeds op gewezen dat van de dimensies die het belangrijkste gevonden worden, ook de verwachtingen vaak het hoogst gespannen zijn (Zeithaml et al. 1990). De kleinste gap bestaat voor de dimensie tastbaarheden (0.31).

Tabel 3. Verwachtingen, percepties en kwaliteitsgaps bij zeven organisaties.

	Gemiddelde			Minimum		Maximum	
	Verw.	Perc.	Kwaliteit (Gap 5)	Verw.	Perc.	Verw.	Perc.
<i>Dimensies</i>							
Tastbare zaken	6.14	5.83	.31	6.01	4.86	6.42	6.19
Betrouwbaarheid	6.42	5.74	.68	6.10	4.91	6.63	6.09
Responsiviteit	6.25	5.82	.43	5.61	5.31	6.61	6.44
Deskundigheid	6.41	5.94	.47	6.04	5.36	6.73	6.31
Empathie	6.14	5.71	.43	5.74	4.92	6.70	6.26
Totaal (ongewogen)	6.27	5.81	.46	5.95	5.11	6.59	6.16
Totaal (gewogen)	6.31	5.77	.54	6.00	4.91	6.63	6.10

Naast de ongewogen gap is de gewogen kwaliteitsgap berekend door rekening te houden met de gewichten van de dimensies. Uit de vergelijking van beide blijkt dat de gewogen gap voor alle organisaties groter is dan de ongewogen gap. Dit betekent dat organisaties relatief het slechtst presteren op de meest belangrijke dimensies. Zeithaml et al. (1990, p. 30) rapporteren eenzelfde bevinding.

Een laatste observatie is dat voor alle organisaties, voor alle dimensies, de verwachtingen hoger zijn dan de perceptie. In deze zin meet SERVQUAL niet zozeer kwaliteit maar vooral 'non-kwaliteit' en ontevredenheid.

4.3 Bruikbaarheid van de meting van de verwachtingen

De verwachtingen worden in SERVQUAL gemeten door de respondent per item aan te laten geven, op een zevenpuntsschaal, wat men verwacht van een *excellente* onderneming in de betreffende branche. Een belangrijke vraag die nu gesteld kan worden is: in hoeverre is het noodzakelijk om verwachtingen van klanten te meten en is de wijze waarop dat in SERVQUAL plaats vindt de meest valide en bruikbare?

Uit praktisch oogpunt is deze vraag relevant omdat het meten van de verwachtingen het aantal vragen aan klanten verdubbelt, hetgeen de non-respons sterk kan verhogen. Ook vanuit theoretisch oogpunt is deze vraag interessant omdat het de kern van de definitie van verwachtingen raakt. Belangrijk bij die definitie is, wat de klant als referentiekader kiest bij het beoordelen van kwaliteit. Zo maakt het nogal verschil of de klant zijn verwachtingen over de 'ideale' kwaliteit (de kwaliteit die de klant 'zou behoren' te krijgen) kiest als referentiekader bij het beoordelen van de dienstverlening, of dat hij de verwachtingen hanteert die hij redelijkerwijs kan verwachten gezien de prijs van de dienstverlening en de overige omstandigheden (Zeithaml et al. 1993, Boulding et al. 1993; zie ook Lemmink 1992). Daarnaast wordt wel gesteld dat de definitie van kwaliteit als verschil tussen verwachtingen en perceptie minder valide is en dat kwaliteit beter opgevat kan worden als een attitude. In die visie kan worden volstaan met het meten van de perceptie (Cronin & Taylor 1992).

Het belangrijkste gevolg van de huidige meetprocedure is dat de verwachtingen op alle items erg hoog zijn zodat er relatief weinig differentiatie ontstaat, noch tussen dimen-

sies, noch tussen respondenten. Het opsporen van segmenten op basis van deze verwachtingen is daardoor vrijwel onmogelijk. Een oorzaak voor het probleem van de hoge scores is waarschijnlijk dat bij het meten van de verwachtingen de respondenten de neiging zullen hebben ieder van de 22 items apart te beschouwen en dan vrijwel altijd een zes of een zeven te scoren op dat item (op de schaal van 1 t/m 7). Met andere woorden, respondenten gaan de 22 items niet onderling vergelijken en zich dan afvragen van welke items ze relatief veel verwachten en op welke items hun verwachtingen duidelijk lager zijn. Deze antwoordneiging wordt nog versterkt door het referentiekader bij de *excellente* onderneming te leggen (een excellente onderneming behoort immers in alle details te excelleren) en een eindige zevenpuntsschaal te gebruiken. Voor een respondent is het dan bijna onmogelijk lage verwachtingen te hebben op een item. Het gevolg is dat een aantal respondenten deze vraag naar verwachtingen als vervelend en nutteloos ervaart, omdat op ieder item hoog gescoord wordt.³

Ter verdediging van de meetprocedure kan opgemerkt worden dat wanneer naar de resultaten bij de zeven organisaties gekeken wordt er, ondanks de meetproblemen, wel degelijk variatie gevonden wordt tussen de organisaties (zie minimum en maximum waarden van de verwachtingen in tabel 3). Tevens blijkt deze variatie zinvol te interpreteren te zijn en overeen te komen met wat men zou veronderstellen. Zo hebben patiënten zeer hoge verwachtingen over de empathie (en ook deskundigheid) in een excellent ziekenhuis, terwijl klanten bij een luchthaven en een restaurant relatief weinig empathie verwachten. Aldaar verwacht men weer veel van een stipte en prompte dienstverlening (responsiviteit). Uit dergelijke resultaten blijkt dat de meting van verwachting in zekere mate valide is.

Het antwoord op de vraag of voor deze zeven organisaties de meting van de verwachtingen nu veel extra en relevante informatie heeft opgeleverd boven alleen de meting van de perceptie, is moeilijker te geven. Als de grootte van gap 5 (verwachting minus perceptie) vergeleken wordt met de grootte van gap 5A (de maximum score 7 minus de perceptie) dan kunnen we afleiden dat wellicht drie van de zeven organisaties zonder de meting van verwachtingen, tot andere conclusies over hun kwaliteit gekomen zouden zijn. Zo heeft de vereniging een kleine gap 5 maar een zeer grote gap 5A. Dit komt omdat de perceptie van de dienstverlening vrij laag is (er is veel ruimte ter verbetering: gap 5A), maar de verwachtingen zijn ook vrij laag gespannen. Iets vergelijkbaars geldt voor de restaurantketen. Voor het ziekenhuis daarentegen geldt het omgekeerde. Daar zijn de verwachtingen zeer hoog gespannen en ondanks de goede beoordeling van de kwaliteit is de kwaliteitgap 5 nog vrij groot. De conclusie is dat als de verwachtingen niet gemeten zouden zijn, het ziekenhuis onterecht 'tevreden' had kunnen zijn, terwijl de vereniging en de restaurantketen zich ten onrechte ongerust zouden hebben gemaakt. Wanneer er meer verschillen tussen de zeven organisaties zouden zijn geweest, in de zin van een aantal A-merken tegenover B-merken, dan zou de waarde van het meten van zowel verwachtingen als percepties waarschijnlijk nog iets duidelijker zijn. Het onderzoek toont overigens tegelijk aan dat het voor de overige vier organisaties weinig extra informatie heeft opgeleverd.

Alles overwegende, leiden de bevindingen ons tot de volgende aanbeveling. In ieder geval dient de *perceptie* van de dienstverlening gemeten te worden. Dit geeft op item- en dimensieniveau essentiële informatie over het oordeel van de klanten over de organisatie. Het meten van de verwachtingen over de excellente onderneming in de branche lijkt relatief weinig extra informatie op te leveren voor de individuele onderneming. Dit

weegt waarschijnlijk niet op tegen de negatieve invloed op de motivatie van de respondent. Het lijkt relevanter om, *naast* de meting van de perceptie, de kwaliteit rechtstreeks te meten op itemniveau. Dit zou bijvoorbeeld als volgt kunnen: 'De vakantiebungalows van X zijn: (1) 'beduidend comfortabeler dan ik verwachtte' t/m (7) 'beduidend minder comfortabel dan ik verwachtte'. Bij de instructie of in de formulering van de antwoordschaal kan aangegeven worden welk referentiekader de respondent bij het beantwoorden in ogenschouw moet nemen. Relevante referentiekaders kunnen zijn: 'een excellent bungalowpark', 'een doorsnee bungalowpark', 'dit bungalowpark X'.

4.4 De betrouwbaarheid van de dimensies

Boven is reeds gesteld dat rapportage van de resultaten op itemniveau een organisatie veel detailinformatie oplevert. Voor het algemene overzicht worden echter de vijf kwaliteitsdimensies gehanteerd. Hierbij wordt verondersteld dat de items op betrouwbare wijze een dimensie samenstellen. Een aantal auteurs heeft dit aspect onderzocht door op de set van items (confirmatorische) factoranalyse uit te voeren (Carman 1990, Saleh & Ryan 1991, Cronin & Taylor 1992, Vogels et al. 1989) met als enigszins teleurstellende conclusie dat de dimensies in aantal en samenstelling niet altijd teruggevonden worden, zoals wordt verondersteld in het model. Lemmink (1992) adviseert dan ook voor elke nieuwe toepassing de dimensiestructuur te onderzoeken.⁴

In zijn algemeenheid is het een goed advies om, als de organisatie tenminste beschikt over voldoende expertise om de factorstructuur van de data te onderzoeken, te besluiten tot het bepalen van de 'organisatiespecifieke' dimensies. Wanneer een organisatie deze expertise niet in huis heeft, is de vraag relevant of een indicatie van het risico gegeven kan worden dat men tot onjuiste conclusies komt als men er vanuit gaat dat de vijf dimensies inderdaad bestaan, zonder dit te checken.

Deze laatste vraag zullen we onderzoeken door de betrouwbaarheid (interne consistentie) van de dimensies te meten met Cronbach's α (Churchill 1990, p.498). De coëfficiënt α ligt tussen 0 (volledig onbetrouwbaar, de items die de dimensie meten hangen niet met elkaar samen) en 1 (volledig betrouwbaar, de items correleren perfect).⁵ Als de items, bijvoorbeeld de vijf items die empathie meten, laag correleren dan betekent dit dat de items duidelijk verschillende aspecten meten en derhalve niet samen de dimensie empathie vormen. Er wordt gestreefd naar een zo hoog mogelijke α waarbij een α van 0.70 een acceptabel niveau is (Nunnally 1978). Als het nu zo is dat voor de zeven organisaties de α 's voldoende hoog zijn, dan kan daaruit afgeleid worden dat de dimensies voldoende consistent gemeten worden door de items. Er is dan weinig reden om via factoranalyse de dimensies te onderzoeken en de conclusies kunnen gebaseerd worden op de vijf dimensies.

In tabel 4 staan de minimum, maximum en mediane⁶ waarden van α per dimensie vermeld. Deze zijn berekend voor de perceptiescores; de α 's berekend op de verwachtingenscores geven een vergelijkbaar beeld. Het blijkt dat voor iedere dimensie de mediane waarden zeer aanvaardbaar zijn (rond de 0.80). Zelfs de minimum waarden zijn vrij goed te noemen: rond de 0.70. Verder lopen de α 's weinig uiteen tussen de organisaties. Hoge waarden zijn gevonden bij het ziekenhuis, relatief minder betrouwbaar zijn de dimensies gemeten bij de touroperator en een bungalowpark. De mediane waarden in ons onderzoek zijn slechts iets lager dan de mediane waarden die Parasuraman et al. (1990) rapporteren voor vijf organisaties (twee banken, twee verzekeringsmaatschappijen en een telecommunicatiebedrijf).

Tabel 4. *Betrouwbaarheid (Cronbach α) per SERVQUAL-dimensie voor zeven dienstverlenende organisaties (percepties, klanten).*

	Aantal items	Mediaan	Minimum	Maximum	Parasuraman (1990)
Tastbare zaken	4	.79	.68	.86	.84
Betrouwbaarheid	5	.87	.73	.90	.90
Responsiviteit	4	.81	.66	.91	.91
Deskundigheid	4	.78	.65	.87	.89
Empathie	5	.79	.72	.87	.87

Op basis van onze resultaten concluderen we dat de vijf kwaliteitsdimensies voldoende betrouwbaar gemeten worden. Het geeft een indicatie van de generaliteit van de 22 items in de vragenlijst en het lijkt een uitvloeisel van de degelijkheid waarmee SERVQUAL is ontwikkeld. Als een organisatie daarentegen besluit veel extra kwaliteitsattributen toe te voegen of formuleringen van items sterk te wijzigen, dan neemt de noodzaak tot een zorgvuldige analyse van de dimensiestructuur duidelijk toe.

5. DE INTERNE KWALITEITSGAPS

5.1 Hoe goed is het management op de hoogte van de verwachtingen van de klanten?

De eerste interne gap in het model betreft het verschil tussen wat klanten verwachten van een excellente organisatie en wat het management denkt dat klanten verwachten. Het kan zijn dat het management niet goed op de hoogte is van wat klanten wensen en daardoor de verwachtingen van klanten over- of onderschat, wat kan leiden tot verkeerde beleidsbeslissingen. Voor vier van de zeven organisaties is de gap negatief: het management onderschat enigszins de verwachtingen van klanten; bij de overige drie organisaties overschat het management de verwachtingen.

Bij zes van de zeven organisaties is deze eerste gap in absolute zin niet erg groot; zo bedraagt het gemiddelde absolute verschil 0.18. Dit betekent dat het management de verwachtingen van klanten vrij goed weet in te schatten. De grote uitzondering is het ziekenhuis waar de gap voor vrijwel alle dimensies ongeveer 2 bedraagt. Dit betekent dat wanneer klanten aangeven een 6.5 te verwachten voor bijvoorbeeld responsiviteit, het management deze verwachting inschat met een 4.5. Het management denkt derhalve dat klanten aanzienlijk *lagere* verwachtingen hebben dan zij in werkelijkheid hebben. Een verklaring voor deze extreme afwijking is niet gemakkelijk te geven. Mogelijk dat er sprake is van een methodologisch probleem in de zin dat het ziekenhuismanagement de vraag wellicht vanuit het eigen perspectief heeft beantwoord in plaats vanuit het klantenperspectief. Men vindt dat klanten niet teveel mogen verwachten, gezien de financiële problemen in de gezondheidszorg. Er zijn echter ook enige aanwijzingen dat het management inderdaad slecht op de hoogte is van de verwachtingen van klanten. Op de drie antecedente factoren (zie figuur 2) die de grootte van gap 1 verklaren scoort het ziekenhuis namelijk ook vrij extreem: vergeleken met de andere organisaties rapporteert men een zeer geringe oriëntatie op marktonderzoek en

is men het zeer sterk eens met de stelling dat er teveel lagen zitten tussen werkvloer en top in de organisatie. Tevens constateert men een gebrek aan 'upward communication'. Ook de touroperator en de vereniging onderschatten systematisch, dat wil zeggen voor alle vijf dimensies, de verwachtingen van klanten. Daarentegen overschat het management van de luchthaven en restaurantketen systematisch de verwachtingen van klanten. De organisaties blijken erg uiteen te lopen wat betreft de antecedente factoren. In het algemeen kan zowel het gebruik van marktonderzoek als de 'upward communication' volgens de meeste managers sterk verbeteren (median score is 4.6 op een zevenpuntschaal). Met name bij een bungalowpark, het ziekenhuis en de restaurantketen is men het sterk eens met de stelling dat er teveel managementniveaus aanwezig zijn. Het management van de touroperator, het andere bungalowpark en de vereniging vindt juist dat de organisatie vrij plat is. Volgens verwachting rapporteren de niet-op-winstgerichte organisaties, het ziekenhuis en de vereniging, de minste aandacht voor marktonderzoek. Het restaurant en de bungalowparken zeggen hierin erg actief te zijn.

5.2 De kwaliteitsgaps: twee, drie en vier

De eerste en de vijfde gap zijn boven reeds besproken. Beide gaps zijn vrij klein vergeleken met de gaps 2, 3 en 4 (zie tabel 5). De mediaan voor deze laatste drie gaps bedraagt 2 punten op een zevenpuntschaal. Deze grootte betekent dat het management en de medewerkers van de organisaties aangeven dat er voor de drie gaps nog veel ruimte is ter verbetering. De grootte van deze gaps komt overigens opnieuw goed overeen met resultaten die Zeithaml et al. (1990) rapporteren.

De vertaling van verwachtingen naar specificaties (gap 2)

De tweede gap betreft de mate waarin het management er in zegt te slagen de dienstverlening die de klanten verwachten te vertalen naar specificaties en standaarden. In het model gaat het daarbij om de mate waarin specificaties formeel zijn vastgelegd, expliciet zijn gemaakt en als zodanig ook worden gecommuniceerd naar de medewerkers. Dit betreft bijvoorbeeld normen met betrekking tot telefonische bereikbaarheid: 'max. ... % van de tijd mag een klant een ingesprekstoon krijgen', of 'een monteur moet binnen ... minuten na de melding op de plek gearriveerd zijn'. Een grote gap zal het gevolg zijn als de specificaties niet-formeel en impliciet zijn: 'we veronderstellen dat de medewerkers zelf wel begrijpen wat wij en de klanten van ze verwachten'.

Van de zeven organisaties blijkt deze gap bij de vereniging en het ziekenhuis het grootst te zijn. Het management hiervan zegt het minst te beschikken over formele regels en richtlijnen voor het niveau van dienstverlening aan klanten. De luchthaven en de touroperator rapporteren de kleinste gap en hebben naar eigen zeggen relatief veel expliciete specificaties.

Als gekeken wordt naar de vijf dimensies dan blijkt dat, zoals verwacht mag worden, de kleinste gap bestaat voor de tastbare zaken (mediaan 1.4). Voor deze dimensie zijn de meest formele specificaties opgesteld. Ook is de gap bij responsiviteit relatief klein (mediaan 1.8). Beide dimensies zijn voor een dienstverlenende organisatie waarschijnlijk ook het eenvoudigst om in specificaties en richtlijnen te vertalen. De grootste gap bestaat bij de empathie (mediaan 2.8). Behalve dat inlevingsvermogen en individuele aandacht voor de klant moeilijk in specificaties te vangen is, is het ook de vraag of formele, expliciete regels hiervoor opgesteld dienen te worden. Het gaat er immers om dat een medewerker tijdens de interactie met de klant deze klant niet als een nummer

beschouwt maar als een uniek exemplaar, waarbij geprobeerd wordt aan de specifieke wensen zoveel mogelijk tegemoet te komen ('een uitzondering op de regel kunnen maken vanwege'). Deze uniciteit in de interactie staat enigszins op gespannen voet met de formaliteit van richtlijnen. Al met al komt de richting van verschillen tussen organisaties en tussen dimensies overeen met wat men mag veronderstellen. Dit bevestigt de validiteit van de meting van deze gap.

Zonder hier in detail op de antecedenten voor deze gap in te gaan, komt uit het onderzoek naar voren dat het management van de restaurantketen ruimschoots het hoogste scoort op het definiëren van kwaliteitsdoelen en het communiceren daarvan ('goal setting'). Het verenigingsmanagement rapporteert een relatief geringe betrokkenheid van zichzelf bij kwaliteit, het stelt weinig expliciete kwaliteitsdoelen op en er bestaat weinig standaardisatie in taken. Opvallend was verder dat het management van het ene bungalowpark op alle vier de antecedente factoren zichzelf positiever beoordeelde dan het management van het andere bungalowpark. Het zou interessant zijn om na te gaan of dit verschil ook in werkelijkheid bestaat of dat het te maken heeft met een verschil in de hoeveelheid zelfkritiek.

Tabel 5. Grootte van de vijf gaps bij zeven organisaties (een grotere score betekent een grotere gap).

Gap	Omschrijving	Mediaan	Min.	Max.
1. Mngt. -/ Klanten	Inschatting van verwachtingen van klanten door het management	.18	.01	1.87
2. Mngt.	Vertaling van verwachtingen naar specificaties	2.09	1.34	4.18
3. Medew.	Uitvoering van diensten volgens de specificaties	1.98	1.21	2.49
4. Medew.	Aansluiting van externe communicatie met uitvoering	2.01	1.17	2.46
5. Klanten	Kwaliteit: verwachtingen minus perceptie	.33	.27	1.01

Het consistent leveren van dienstverlening volgens de specificaties (gap 3)

De derde gap betreft de mate waarin de medewerkers er in slagen de, formele dan wel informele, richtlijnen consistent te vertalen naar de uitvoering. Met andere woorden, achten de medewerkers zich in staat vrijwel altijd de dienst op het gewenste niveau te verlenen?

De verschillen tussen de zeven organisaties zijn bij deze gap minder groot dan bij de tweede gap (vergelijk de minimum en maximum waarden in tabel 5). Medewerkers bij één van de bungalowparken en bij de vereniging rapporteren de grootste gap. Bij het bungalowpark is het vooral de dimensie betrouwbaarheid ('doen wat je belooft') die volgens medewerkers nog weleens in de problemen komt. Bij de vereniging betreft het drie dimensies: tastbare zaken, betrouwbaarheid en empathie. Het restaurantpersoneel

ziet de minste problemen in het produceren van de dienst volgens de specificaties. Wanneer de derde gap bekeken wordt naar dimensies, dan blijkt opnieuw de dimensie empathie de meeste problemen op te leveren voor de zeven organisaties. Blijkbaar vinden medewerkers het moeilijk de klant voldoende individuele aandacht te geven. In het bijzonder zal dit bijvoorbeeld gelden tijdens piektijden. Zo is er vaak een trade-off nodig tussen het snel wegwerken van een wachtrij en de ruime aandacht voor de klant die aan de beurt is. De dienstverlener ziet de rij ongeduldigen, maar wil toch de huidige klant hoge kwaliteit leveren. Omgekeerd wil iedere klant die in de rij staat dat de dienstverlener opschiet bij degenen die voor hem/haar staan, maar ruim de tijd neemt als men zelf aan de beurt is.

Bij mogelijke oorzaken van een grote gap kan gedacht worden aan aspecten zoals: de juiste persoon zit niet op de juiste plaats, de medewerker krijgt te weinig faciliteiten van de organisatie om het werk naar behoren te doen, het niet beloond worden voor extra inspanning op kwaliteitsgebied (zie figuur 2). Slechts enkele resultaten kunnen hier vermeld worden. Typische scores op deze antecedenen liggen tussen de 4 en 5. Een uitzondering is de beoordeling van het teamwork dat in alle organisaties vrij hoog scoort (mediaan 5.6 op een schaal van 7). De verschillen tussen de organisaties in teamwork zijn klein, alhoewel de medewerkers op de luchthaven zich het minste een deel van een team voelen. De 'juiste man op de juiste plaats' lijkt de meeste problemen op te leveren in één van de bungalowparken, terwijl de medewerkers bij de vereniging en de luchthaven het minst te spreken zijn over de faciliteiten die zij krijgen om hun werk naar behoren te verrichten. Medewerkers bij de touroperator en de restaurantketen zijn over dit aspect het best te spreken.

Het consistent kunnen voldoen aan beloftes (gap 4)

De vierde gap heeft betrekking op het verschil tussen de beloftes die gedaan worden aan de klant in de externe communicatie en de werkelijke dienstverlening die geboden wordt. Om te kunnen voldoen aan de beloften dienen medewerkers ook tijdig op de hoogte zijn van het aanbod en de inhoud van nieuwe diensten, bijvoorbeeld een plotse-ling ingeplande promotieactie in vliegreizen. Een goede horizontale communicatie is daarvoor vereist. Zijn de medewerkers op de hoogte van de beloften en kunnen zij naar eigen zeggen daaraan consistent tegemoet komen?

De mediane grootte van deze gap bedraagt 2 punten op een zevenpuntsschaal zodat ook bij dit aspect er nog veel ruimte is ter verbetering. De minste moeite om tegemoet te komen aan de beloftes rapporteren de medewerkers bij de touroperator (1.2). Ook het luchthavenpersoneel en het restaurantpersoneel ziet hier weinig problemen. De medewerkers van de vereniging hebben relatief de meeste moeite met deze gap (2.5), gevolgd door het ziekenhuispersoneel (2.4). Mogelijk komt dit bij deze organisaties vooral door het ontbreken van duidelijkheid over wat er precies beloofd wordt aan de klant, zodat men niet weet waaraan men als medewerker tegemoet moet komen. Immers, beide niet-op-winstgerichte organisaties beschikken over minder brochures, reclamecampagnes en overige externe communicatie om het imago gestalte te geven dan de andere organisaties. Daarnaast geldt in het bijzonder voor het ziekenhuis dat vooral de vrije publiciteit het verwachtingsniveau van de klanten bepaalt. Het ziekenhuispersoneel geeft zeer duidelijk aan dat zij vinden dat er onredelijke verwachtingen van de gezondheidszorg geschapen worden (de 'propensity to overpromise' scoort zeer hoog). De medewerkers van de vereniging klagen meer over het gebrek aan horizontale communicatie in hun organisatie.

Wat de dimensies betreft komt de betrouwbaarheid als lastigste dimensie naar voren. Dit is niet zo verwonderlijk aangezien deze dimensie ook inhoudelijk het meeste te maken heeft met 'doen wat je belooft'. De medewerkers bij de bungalowparken rapporteren de meeste moeite bij deze gap te hebben met betrekking tot de tastbare zaken. Zij staan voor de taak 'de mooie plaatjes' in brochures waar te maken.

Bruikbaarheid van de drie gaps voor het beleid

De gaps 2 t/m 4 worden in SERVQUAL op dimensieniveau gemeten. Voor ieder van de vijf dimensies moeten het management (voor gap 2) of de medewerkers (voor gaps 3 en 4) aangeven in hoeverre er bijvoorbeeld formele specificaties zijn voor betrouwbaarheid, empathie, etc. Deze metingen zijn daarom van een hoog abstractieniveau waardoor met name een aantal medewerkers duidelijk moeite heeft met het formuleren van een zinvol antwoord op deze vragen.⁷ Echter ook de interpretatie van de gevonden gaps en de vertaling van de resultaten naar concrete aanbevelingen voor het beleid zijn lastig. Een belangrijke vraag is daarom of de huidige meting van de gaps veel informatie oplevert voor het beleid. Deze vraag is tijdens het SERVQUAL onderzoek bij de zeven organisaties vaak aan de orde geweest. Meestal werd de vraag helaas negatief beantwoord.

Als alternatief voor de huidige meting stellen we voor om te overwegen de gaps 2 t/m 4 telkens te meten voor alle 22 items waarop de kwaliteit beoordeeld wordt. Behalve dat de vragen hierdoor eenvoudiger te beantwoorden zijn hetgeen de validiteit en betrouwbaarheid van de informatie ten goede zal komen, geeft dit ook meer precieze informatie over waar volgens medewerkers of management kwaliteitsproblemen aanwezig zijn bij de dienstverlening. Op deze wijze wordt duidelijker bij welke aspecten er specificaties ontbreken of op welke aspecten medewerkers vinden dat geleverde dienst niet voldoet aan de gestelde specificaties. Tevens kunnen op deze wijze de verschillen in inzicht tussen de respondenten gemakkelijker geïnterpreteerd en benut worden. Het is erg belangrijk om na te kunnen gaan over welke van de 22 items het management onderling sterk van mening verschilt en over welke aspecten eensgezindheid bestaat. Uiteraard wordt de vragenlijst door dit voorstel een stuk langer. Wij denken echter dat een organisatie die de mening van het management en medewerkers ziet als een essentiële bron van informatie over het gevoerde kwaliteitsbeleid, in staat mag worden geacht de respondenten voldoende te kunnen motiveren tot het invullen van de vragenlijsten. Het leveren van kwaliteit in dienstverlening buiten de eigen mensen om is nu eenmaal onmogelijk. Het is de bedoeling het bovenstaande voorstel binnenkort op uitvoerbaarheid en validiteit te testen.

5.4 Antecedenten voor de gaps

Voor ieder van de vier interne gaps zijn in het SERVQUAL-model antecedente factoren geselecteerd die de grootte van de gap zouden verklaren (zie figuur 2). In deze paragraaf zal nagegaan worden of deze antecedenten betrouwbaar gemeten worden in de huidige vragenlijst. Er moet hierbij gesteld worden dat de auteurs zelf al aangeven dat voor dit deel van het model nog veel werk verzet moet worden ter validering van de antecedenten en de bijbehorende vragenlijst. In tabel 6 worden per gap deze antecedenten opgesomd tesamen met de Cronbach α 's die in ons onderzoek gevonden zijn. Deze worden vergeleken met de α 's die Parasuraman et al. (1990) vermelden. Overigens zijn

twee antecedenten ('levels of management' en 'technology-job fit') gemeten met slechts één item, zodat hierbij geen α berekend kan worden.

Tabel 6 geeft duidelijk aan dat voor de meeste antecedenten de mediane waarde van de zodanig laag is dat niet gesproken kan worden van een betrouwbaar gemeten antecedent. Dit betekent dat de items te veel op zichzelf staan. Soms zijn er zelfs negatieve correlaties tussen items die dezelfde antecedente factor veronderstellen te meten. Het patroon en de orde van grootte van de α 's in ons onderzoek komen overigens verrassend goed overeen met de waarden die Parasuraman et al. (1990) rapporteren.

Tabel 6. Betrouwbaarheid (Cronbach α) voor de antecedenten die de gaps beïnvloeden, gemeten bij zeven organisaties.

Gap	Antecedent	Aantal items	Mediaan	Min.	Max.	Parasu. (1990)
1.	Marketing research orientation	4	.50	.36	.65	.60
	Upward communication	4	.60	.16	.90	.56
	Levels of management	1	-	-	-	-
2.	Mngt. commitment to quality	4	.42	.05	.73	.54
	Goal setting	2	.40	< 0	.87	.84
	Task standardization	2	.75	< 0	.89	.69
	Perception of feasibility	3	.29	.04	.44	.59
3.	Teamwork	5	.68	.56	.84	.80
	Employee-job fit	2	.56	.30	.68	.34
	Technology-job fit	1	-	-	-	-
	Perceived control	4	.43	.23	.61	.34
	Supervisory control systems	3	.23	< 0	.59	.42
	Role conflict	4	.45	.23	.72	.61
	Role ambiguity	5	.49	.01	.69	.60
4.	Horizontal communication	4	.19	< 0	.56	.49
	Propensity to overpromise	2	.07	.29	.64	.28

De α 's laten tussen antecedenten en tussen organisaties grote verschillen zien. Een aantal antecedenten wordt redelijk betrouwbaar gemeten zoals 'task standardization', 'teamwork' en de 'upward communication'. Maar zelfs bij deze antecedenten zijn de verschillen tussen de organisaties groot zoals de minimum en maximum waarden laten zien. Erg onbetrouwbaar zijn de antecedenten van gap 4: de horizontale communicatie en de neiging om teveel te beloven. De maximum waarden van de α zijn daarbij uitschieters. Ook de 'perception of feasibility' (gap 2) wordt bij alle organisaties onbetrouwbaar gemeten. Er is tevens gekeken of de betrouwbaarheid bij bepaalde organisaties systematisch hoger is dan bij andere organisaties. Hieruit kwam geen duidelijk patroon naar voren. Er was een lichte tendentie dat bij de vereniging, de luchthaven en het ziekenhuis, de antecedenten iets betrouwbaarder zijn gemeten dan bij de andere organisaties. De verschillen zijn echter erg klein en zijn, gegeven de lage absolute waarden van de α 's, niet van grote betekenis.

De volgende conclusies kunnen op basis van tabel 6 getrokken worden. Ten eerste kan een organisatie niet zonder meer somscores over items berekenen en dan in de interpretatie van SERVQUAL de benaming van de antecedenten gebruiken. Immers, als bijvoorbeeld vier items die verondersteld worden samen de horizontale communicatie in het bedrijf te meten nauwelijks samenhangen, dan worden als het ware vier losse aspecten van communicatie gemeten. Het is dan het beste om de vier items individueel te beschouwen en daar de interpretatie op te richten. Ten tweede is duidelijk dat verder wetenschappelijk onderzoek dringend noodzakelijk is naar juist deze antecedenten. Niet alleen naar de juistheid en volledigheid van de antecedenten zal gekeken moeten worden, maar ook naar de manier waarop deze gemeten worden. Andere vragen in de vragenlijst zijn nodig om de interne factoren die een succesvol kwaliteitsbeleid bepalen goed in de vingers te krijgen.

Nader onderzoek is des te meer gewenst omdat de belangstelling van het management voor deze antecedenten groot is. Een opvallend resultaat van de SERVQUAL toepassing was dat juist de antecedenten relatief veel discussie los maakten bij de organisaties. Niet alleen was het vaak de eerste keer dat een aantal interne factoren grootschalig en op een standaardmanier gemeten werden, maar bovendien kreeg het management enig inzicht in hoe verschillend er binnen het management en bij de medewerkers werd gedacht over het reilen en zeilen in de organisatie. Wel is het zo dat SERVQUAL zelf weinig houvast biedt hoe de resultaten geïmplementeerd moeten worden ter verbetering van het kwaliteitsbeleid. Daardoor ontstond soms het idee dat de organisatie na een SERVQUAL-toepassing in de kou staat omdat er wel gaps geconstateerd zijn en er veel discussie is losgemaakt, maar de oplossingen nog niet voor de hand liggen. Wat dit betreft dienen, onterechte, verwachtingen over de methode bijgesteld te worden. SERVQUAL is in eerste instantie bedoeld als diagnosemethode.

Een toepassing van SERVQUAL zal gevolgd moeten worden door maatregelen om gediagnosticeerde problemen op te lossen of aan te pakken. Het invoeren van een integraal kwaliteitssysteem raakt uiteraard de essentiële manieren van werken in een organisatie (zoals die naar voren komen in de antecedente factoren), waar ook anderen dan alleen het marketing management bij betrokken moet worden en wat een continue zorg van de hele organisatie moet zijn. SERVQUAL is daarbij slechts één van de hulpmiddelen.

6. CONCLUSIES EN AANBEVELINGEN

SERVQUAL is een model en met bijbehorende methode dat organisaties kan helpen een diagnose te stellen over waar het goed of juist minder goed gaat met het kwaliteitsbeleid. Zelfs zonder empirisch onderzoek kan het model het management helpen het beleid door te lichten. De methode geeft vooral een brede analyse van het probleem, alsof er een overzichtsfoto wordt genomen van de stand van zaken. Vaak zal het nodig zijn een SERVQUAL-onderzoek aan te vullen met andere methoden om daarmee in te zoomen op gesignaleerde gaps en zodoende diepgaande en gedetailleerde kennis te verkrijgen.

Onze ervaringen met zeven toepassingen laten zien dat het een waardevolle methodiek is die tevens redelijk robuust is. Het kan toegepast worden op zeer verschillende soorten organisaties en geeft daarbij in het algemeen vrij betrouwbare en valide informatie. De risico's om bij een toepassing tot duidelijk onjuiste conclusies te komen, zijn

gering. De resultaten geven ons wel aanleiding tot de volgende kritische kanttekeningen en aanbevelingen.

Aanbevelingen voor het externe deel

- De kwaliteit van dienstverlening wordt in SERVQUAL gemeten met 22 items die gebaseerd zijn op vijf onderliggende dimensies (eigenlijk omvatten deze vijf dimensies tien oorspronkelijk geformuleerde dimensies, zie boven). Bij een toepassing lijkt het raadzaam dat de organisatie nagaat of deze 22 items alle aspecten dekken of dat er extra items nodig zijn om de kwaliteit volledig in kaart te brengen (een 22⁺ vragenlijst). De aanbeveling zou zijn om bij het kiezen van extra, organisatiespecifieke items, uit te gaan van de oorspronkelijke tien kwaliteitsdimensies en op basis van die dimensies items te formuleren. In voorkomende gevallen kan het zijn dat er naast deze tien dimensies nog extra dimensies zijn die gedekt moeten worden door items.
- De huidige lijst met items blijkt met vrij grote betrouwbaarheid de vijf dimensies van kwaliteit te meten. Er zijn geen aanwijzingen dat bij een toepassing van SERVQUAL een uitgebreid onderzoek naar de factorstructuur plaats dient te vinden. Wel is het zo dat wanneer items worden toegevoegd aan de lijst van 22, of als anderszins de lijst sterk gewijzigd wordt, de kans op een nieuwe dimensiestructuur toeneemt, zodat wel extra analyses vereist zijn.
- Bij de meting van de kwaliteit is het van belang in ieder geval de perceptie van de dienstverlening te meten. Het onderzoek laat zien dat de toegevoegde waarde van de meting van de verwachtingen van klanten, zoals dat in het huidige SERVQUAL plaats vindt, gering is. Verder wetenschappelijk onderzoek naar de definitie en meting van verwachtingen is noodzakelijk. Voor de praktijk lijkt het in de tussentijd relevant om, naast de perceptie, voor ieder van 22 items de kwaliteit te meten. Dat kan door per item de klant aan te laten geven of de geleverde dienst boven of onder verwachting bleef.
- Als de meting van de verwachtingen vervangen wordt door een kwaliteitsmeting per item, dan is het niet meer mogelijk gap 5 te meten: de inschatting van het management van de verwachtingen van de klant. Uit het onderzoek blijkt dat deze meting ook weinig extra informatie oplevert. Voor zes van de zeven organisaties bleek de gap erg klein. Als alternatief kan dan de inschatting van het management van de *perceptie* van klanten gemeten worden, of, als tweede alternatief, hen een inschatting te laten geven welke aspecten van de dienstverlening de klanten mee- of tegenvallen.
- Naast de bestaande vragen is het aan te bevelen een aantal metingen van variabelen toe te voegen die de waarde van een SERVQUAL-klantenonderzoek sterk kunnen verhogen zonder substantiële extra kosten. Hierbij wordt gedacht aan: (a) een holistische directe meting van de kwaliteit van dienstverlening,⁸ (b) de prijs/kwaliteitsverhouding, (c) of de klant, alles overziende, vindt dat de dienstverlening mee- of tegenviel (een holistische indirecte kwaliteitsmeting); (d) de waarschijnlijkheid dat men de volgende keer dezelfde dienstverlener kiest, (e) of men de dienst zou aanbevelen bij vrienden en kennissen, en (f) de tevredenheid met de dienstverlening. Deze algemene oordelen kunnen vervolgens verbonden worden met de beoordeling op de aparte items om op die manier een validering van de meting te verkrijgen (zie ook Cronin & Taylor 1992).

Aanbevelingen voor het interne deel

- De interne gaps blijken vrij betrouwbaar gemeten te kunnen worden. Daarnaast zijn de verschillen die met SERVQUAL gevonden worden tussen de organisaties een duidelijke indicatie van de validiteit van de gemeten gaps. Veronderstellingen over de grootte van gaps bij organisaties waarvan gezien de aard van de dienstverlening of het gevoerde marketingbeleid verwacht mocht worden dat een bepaalde gap voor een bepaalde dimensie groter of kleiner zou zijn die dan bij andere organisaties, bleken vaak bevestigd te worden.
- Wel is vastgesteld dat door het meten van de gaps op het niveau van de dimensies, de interpretatie en implementatie van de bevindingen beperkt is. Voorgesteld wordt om de gaps 2 t/m 4 te meten voor ieder van de 22 items. Dit levert niet alleen veel meer gedetailleerde informatie op die meer richtinggevend kan zijn voor het beleid, maar zal tevens de validiteit en betrouwbaarheid van de meting ten goede komen omdat de metingen op deze manier veel concreter en eenduidiger zijn.
- De antecedente factoren voor de gaps blijken niet betrouwbaar gemeten te worden. De vragen waarmee de antecedenten gemeten worden staan teveel op zichzelf om te mogen spreken van factoren. Aanbevolen wordt de informatie over de antecedenten op individueel vraagniveau te rapporteren en niet samen te vatten onder de namen van antecedente factoren zoals verondersteld wordt in SERVQUAL. Deze gedetailleerde informatie over antecedenten blijkt overigens een levendige bron van discussie te zijn en is dan ook erg waardevol. Het zal duidelijk zijn dat nog nader wetenschappelijk onderzoek gewenst is naar de identificatie en meting van de interne succesfactoren voor de externe kwaliteit.

NOTEN

1. De auteur dankt het management en de medewerkers van de zeven organisaties voor hun medewerking aan het onderzoek. Tevens dankt hij de deelnemers aan het vak Marketing van Diensten (MM9), voor de data-verzameling en de initiële analyse van de resultaten. Aan het artikel werd gewerkt tijdens een onderzoeks-verblijf aan The Fuqua School of Business, Duke University, Durham, USA. Dit verblijf werd mogelijk gemaakt door de toekenning van een NWO-stipendium.
2. De dimensie deskundigheid en beleefdheid ('assurance') bestaat op haar beurt uit vier onderliggende dimensies: 'competence', 'courtesy', 'credibility' en 'security'. De dimensie empathie wordt gevormd door drie dimensies: 'access' (benaderbaarheid), 'communication' (spreek de taal van de klant) en 'understanding the customer' (ken de klant "persoonlijk"). In plaats van vijf dimensies wordt daarom ook wel van tien kwaliteitsdimensies gesproken.
3. Men kan zelfs veronderstellen dat sommige respondenten zessen scoren om niet als 'consistent extreem' over te komen. Ter reductie van de problemen kan overwogen worden magnitude scaling (Lodge 1981) te gebruiken als alternatief voor de zevenpuntsschaal. Op deze wijze heeft de respondent onbeperkt ruimte naar boven en kan hij subtiele verschillen beter tot uitdrukking brengen.
4. Voor een deel zijn de gerapporteerde problemen veroorzaakt door de negatief geformuleerde items voor responsiviteit en empathie (Lemmink 1992). In de nieuwste versie van SERVQUAL zijn deze items echter al positief geformuleerd (zie Zeithaml et al. 1990) zodat dit probleem inmiddels ondervangen is. Voor een ander deel is de inconsistentie in factorstructuur te wijten aan het feit dat substantiële veranderingen in de formulering worden aangebracht of dat items verwijderd of toegevoegd worden, waardoor de structuur per definitie verandert en instabieler wordt (Parasuraman et al. 1991).
5. De grootte van α wordt bepaald door het aantal items in de schaal en de gemiddelde correlatie tussen de items. De betrouwbaarheid stijgt als de gemiddelde correlatie stijgt. Naarmate er meer items in de schaal zitten, stijgt de betrouwbaarheid eveneens (ceteris paribus).
6. Ter bepaling van de mediane waarde worden de zeven α 's gerangordend naar grootte, waarna de middelste α (de vierde) wordt geselecteerd.
7. De gaps worden overigens wel vrij betrouwbaar gemeten door de vijf dimensies. De mediane Cronbach α

- bedraagt voor gap 2 0.72, voor gap 3 0.73 en voor gap 4 0.81. Deze waarden zijn slechts iets lager dan de α 's die Parasuraman et al. (1990) rapporteren voor de gaps. De α 's geven aan dat de kwaliteitsdimensies niet onafhankelijk van elkaar zijn, maar onderling samenhangen.
8. De vraag zou kunnen luiden: 'Als u nu alle aspecten op een rijtje zet, wat vindt u dan van de kwaliteit van de dienstverlening door X?' met als antwoordschaal: (1) 'heel lage kwaliteit' t/m (7) 'heel hoge kwaliteit'. Op hetzelfde algemene niveau kunnen op zevenpuntsschalen de overige variabelen gemeten worden.

LITERATUUR

- Boulding, W., A. Kalra, R. Staelin and V.A. Zeithaml, 1993. A dynamic process model of service quality: From expectations to behavioral intentions. *Journal of Marketing Research*, 30, 7-27.
- Carman, J.M., 1990. Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions. *Journal of Retailing*, 66, 33-55.
- Churchill, G.A., 1990. *Marketing research: Methodological foundations*. Chicago: The Dryden Press.
- Cronin, J.J. and S.A. Taylor, 1992. Measuring service quality: A reexamination and extension. *Journal of Marketing*, 56, 55-68.
- Lemmink, J.G.A.M., 1992. Metingen ten behoeve van kwaliteitsmanagement: een kritische beschouwing van het SERVQUAL-model. In: A.E. Bronner et al. (red.), *Jaarboek van de Nederlandse Vereniging van Marktonderzoekers '92-'93*. Haarlem: De Vrieseborch (pp. 223-235).
- Lodge, M., 1981. *Magnitude scaling: Quantitative measurement of opinions*. Beverly Hills: Sage.
- Nunnally, J.C., 1978. *Psychometric theory* (2nd ed.). New York: McGraw Hill.
- Parasuraman, A., V.A. Zeithaml and L.L. Berry, 1985. A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49, 41-50.
- Parasuraman, A., V.A. Zeithaml and L.L. Berry, 1988. SERVQUAL: A multiple item scale for measuring customer perceptions of service quality. *Journal of Retailing*, 64, 12-40.
- Parasuraman, A., L.L. Berry and V.A. Zeithaml, 1990. *An empirical examination of relationships in an extended service quality model*. Cambridge: Marketing Science Institute; Report no. 90-112.
- Parasuraman, A., L.L. Berry and V.A. Zeithaml, 1991. *Refinement and reassessment of the SERVQUAL scale*. Working paper The Fuqua School of Business, Duke University.
- Saleh, F. and C. Ryan, 1991. Analysing service quality in the hospitality industry using the SERVQUAL model. *The Services Industries Journal*, 11, 324-343.
- Vogels, R.J.M., J.G.A.M. Lemmink en J.D.P. Kasper, 1989. Some methodological remarks on the SERVQUAL model. In: G.J. Avlonitis et al. (eds.), *Marketing thought and practice in the 1990's*, Proceedings of the 27th EMAC conference, Athens (pp. 789-800).
- Zeithaml, V.A., A. Parasuraman and L.L. Berry, 1990. *Delivering service quality*. New York: The Free Press.
- Zeithaml, V.A., L.L. Berry and A. Parasuraman, 1993. The nature and determinants of customer expectations of service. *Journal of the Academy of Marketing Science*, 21, 1-12.